Attorney Docket No.: 081862.P101 *PATENT*

UNITED STATES PATENT APPLICATION

FOR

# SYSTEM AND METHOD FOR BANDWIDTH PROTECTION IN A PACKET NETWORK

INVENTOR:

DAVID A. HUGHES

PREPARED BY:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 WILSHIRE BOULEVARD
SEVENTH FLOOR
LOS ANGELES, CA 90025-1026

(408) 720-8300

"Express Mail" mailing label number EL431891232US

Date of Deposit December 23, 1999

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231

CHRISTOPHER P. MARSHALL
(Typed or printed name of person mailing paper or fee)

C.P. Marshall 12/23/99
*(Signature of person mailing paper or fee)* *Date*

# SYSTEM AND METHOD FOR BANDWIDTH PROTECTION IN A PACKET NETWORK

## FIELD OF THE INVENTION

The present invention relates generally to telecommunication systems, and, more particularly, to a system and method for providing guaranteed bandwidth protection and for maintaining quality of service in a connection-oriented network, such as an Asynchronous Transfer Mode (ATM) network.

## BACKGROUND OF THE INVENTION

ATM is a switching and multiplexing technique designed for transmitting digital information, such as data, video, and voice, at high speed, with low delay, over a telecommunications network. The ATM network includes a number of switching nodes coupled through communication links. In the ATM network, bandwidth capacity is allocated to fixed-sized units named "cells." The communication links transport the cells from a switching node to another. These communication links can support many virtual connections, also named channels, between the switching nodes. The virtual connections, such as a Virtual Channel Connection (VCC), assure the flow and delivery of information contained in the cells.

Each cell contains a cell header and cell data. The cell header includes information necessary to identify the destination of that cell. The components of the cell header include, among other things, a Virtual Channel Identifier (VCI) and a Virtual Path Identifier (VPI), for collectively identifying an ATM connection for that particular cell, and a Payload Type Identifier (PTI), for indicating whether the cell is sent from one user to another, whether cell data refers to administration or management traffic, and whether congestion is present within the network.

The ATM Forum, which is a user and vendor group establishing ATM standards, has also defined several ATM service categories, used in characterization of a virtual connection. For example, there are 16 quality of service categories, among them categories such as (1) a Constant Bit Rate (CBR), which supports a constant or guaranteed rate to transport services, such as video or voice, as well as circuit emulation, which requires rigorous timing control and performance parameters; (2) a Variable Bit Rate (VBR), real time and non real time, which supports variable bit rate data traffic with average and peak traffic parameters; (3) an Available Bit Rate (ABR), which supports feedback to control the source rate in response to changed characteristics in the network; and (4) an Unspecified Bit Rate (UBR).

Figure 1a illustrates a prior art ATM network 100, typically including several network nodes 110 connected through single communication links 120. Typically, nodes 110 are located in the middle of the network 100, but may also be located at the edges. The network 100 is a data transmission network with guaranteed bandwidth and quality of service. Users 130 access the network 100 and connect to the nodes 110 via similar links 120. Generally, the illustrated communication links support multiple virtual connections. If one link fails, the cells of information transmitted along the virtual connections supported by that link are re-routed using other links in order to avoid extensive loss of data. However, the process of switching and re-routing is time consuming, typically a few seconds, and some loss of data is inevitable.

Attempts made to design a faster switching system for voice communications, preferably under 50 ms, have resulted in the development of an automatic protection switching (APS) system. As illustrated in Figure 1b, in the ATM network 100, parallel links 122, 124 connect the nodes 110 and are used to transmit duplicate information between the nodes 110 and to ensure

fast and reliable data transmission. Link 122 is called an "active" link, while link 124 is a "stand-by" link. Because the same information is transmitted on both links 122, 124, the switching node 110 located at the receiving end can choose either link to receive the transmitted information. For example, if the active link 122 fails, the stand-by link 124 can deliver the same information to the switching node. This APS configuration can be implemented, for example, with the SONET/SDH standards, and can also be used to transport data packets instead of voice communications.

Although the APS system increases the reliability of ATM networks, the duplicated data sent on both links 122, 124 reduces the bandwidth in half, resulting in a waste of bandwidth.

SUMMARY OF THE INVENTION

A system and method are disclosed for protecting a network. The system comprises a plurality of links located in the network. A transmitter switch, coupled to the first end of each link of the plurality of links, is provided for transmitting distinct data along each link. A receiver switch, coupled to the second end of each link, is provided for receiving the distinct data from each link. Distinct data transmitted along one link may be switched to another link when a failure is detected on the one link.

Other features and advantages of the present invention will be apparent from the accompanying drawings, and from the detailed description, which follows below.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

**Figure 1a** shows an exemplary prior art network.

**Figure 1b** shows a prior art network with an example of an Automatic Protection Switching (APS) system.

**Figure 2** is a block diagram of one embodiment of a system with a transmitter switch and a receiver switch, each having one line card coupled to both communication links.

**Figure 3** is a block diagram of the line cards shown in Figure 2.

**Figure 4a** is a flowchart representing the process of assigning virtual connections to appropriate buffer sets.

**Figure 4b** is a flowchart representing the process of storing cells in designated buffer sets.

**Figure 4c** is a flowchart representing the process of selecting a communication link.

**Figure 5** is a block diagram of another embodiment of a system with a transmitter switch and a receiver switch, each having one line card coupled to each communication link.

**Figure 6** is a block diagram of the line cards shown in Figure 4.

DETAILED DESCRIPTION

According to embodiments described herein, the system includes two or more communication links located within a connection-oriented network. A transmitter switch is coupled to one end of the links, while a receiver switch is coupled to the other end. The transmitter switch receives cells containing information, classifies them according to each cell's header information, and stores them within queuing buffer sets. A multiplexer transfers distinct stored cells to each communication link for transmission. By sending distinct data along each link, the system avoids a reduction in bandwidth, experienced in systems which send duplicated data along multiple links.

If a failure is detected on a particular link, the multiplexer can switch the traffic of cells along that link to another link, thereby increasing the reliability of the network. By switching the cells to a fully operational link, the system maintains the guaranteed bandwidth and quality of service for each connection transmitted between the two switches.

The following discussion is presented in the context of an Asynchronous Transfer Mode (ATM) network. It should be understood that the present invention is not limited to ATM networks and may be implemented with other types of networks, such as Frame Relay networks.

Figure 2 illustrates one embodiment of the present invention. A system 200, residing in an ATM network (not shown) comprises a transmitter switch 220 coupled to a receiver switch 240 via communication links 260 and 280. The transmitter switch 220 receives data on input links 210 and 211 and transfers the data to a transmitter line card 300 described in detail below. The transmitter line card 300 is coupled to both communication links 260 and 280 and facilitates the transmission of data along links 260, 280 to the receiver switch 240. The receiver switch 240 includes a receiver line card 360, which will also be described in detail below. The receiver line card 360 is coupled to links 260, 280 and receives data transmitted along those links. Finally, within the receiver switch 240, data is transferred from the receiver line card 360 to output links

290, 291 for further processing. Those skilled in the art will recognize that the exemplary environment illustrated in Figure 2 is not intended to limit the present invention. Indeed, those skilled in the art will recognize that other alternative hardware environments may be used without departing from the scope of the present invention. For example, in one embodiment, the system is a full duplex design, wherein a single line card includes both transmitting and receiving modes. The transmitter line card 300 and the receiver line card 360 will now be described in further detail.

According to Figure 3, the transmitter line card 300 receives data from input links 210 and 211. Data received by line card 300 includes individual cells of information, which are transmitted through virtual connections (VCs) established in the ATM network. It is known in the art that each communication link in the ATM network may carry many different VCs. The line card 300 further comprises a classifier device 310, two sets of queuing buffers 320 and 330, a multiplexing device 340, and a processor 350.

Each queuing buffer set 320, 330 further includes multiple quality of service (QoS) queues, for example a variable bit rate (VBR), a constant bit rate (CBR), an available bit rate (ABR), and an unspecified bit rate (UBR) service queues. In addition, there may be per connection queues. Although the figure only illustrates three quality of service queues, the number of queues in each buffer set may be much greater. Although one embodiment includes two queuing buffer sets for the two links 260, 280, for alternative embodiments multiple queuing buffer sets, corresponding to multiple communication links, may be used. Alternatively, all queuing buffer sets could be stored in a single memory system. The queuing buffer sets 320, 330 are coupled to the multiplexing device 340, which in turn is coupled to the communication links 260 and 280.

The transmitter switch 220 receives a VC setup request, for example from an adjacent node or network management system, and assigns each VC to one of the buffer sets. The switch 220 also programs this information into the

classifier 310. When the classifier 310 receives the cells associated with the VCs from input links 210, 211, it classifies the cells based on each cell's header information, such as their virtual path identifier (VPI) and virtual channel identifier (VCI), and transfers the classified cells to either one of the queuing buffer sets 320, 330, based on the VPI and VCI of each cell. In particular, in order to determine which queuing buffer set to receive each particular cell, the classifier 310 examines the VPI and VCI of each cell and generates a logical connection number (LCN), which correlates the VPI/VCI of each cell with a VCI, VPI, and port used to transfer the cell. The classifier 310 stores all cells associated with a VC, having the same LCN, and correspondingly the same VPI and VCI, to the same queuing buffer set.

Figure 4a is a flowchart representing a detailed process of assigning VCs to appropriate buffer sets. The process of assigning VCs is performed in accordance with a connection admission control (CAC) process described in further detail below.

As shown in the flowchart of Figure 4a, at step 410, the transmitter switch 220 shown in Figure 2 receives a VC request. The request may originate from a user (not shown), from a router processor (not shown) coupled to multiple transmitter switches, from a network operator's operation systems (not shown), or from any other entity within the ATM network. At step 420, the transmitter switch 220 performs a CAC process to determine whether the communication links can accommodate the VC request. In the preferred embodiment, the transmitter switch 220 maintains a CAC process over the link pair 260, 280 to determine whether new data may be transmitted along those links. The CAC process ensures that the QoS objectives of all VCs transmitted on both links 260, 280 can be met, even if a failure of either links is detected. If links 260, 280 are different in size and the larger link fails, the CAC process must ensure that the QoS objectives can be met when all VCs are re-routed along the smaller link. In one specific example, the transmitter switch 220 determines the equivalent bandwidth (EBW) required to meet the QoS

objectives of the requested VC. If the sum of the EBW of all existing VCs on both links 260, 280, including the EBW of the requested connection, is less than the total bandwidth of the smaller link, assuming links 260, 280 are different in size, then the VC request is accepted. Otherwise, the VC request is denied at step 430.

If the VC request is accepted, at the next step 440, the switch 220 must select one of the queuing buffer sets 320 and 330 to assign the VC. For one embodiment, the switch 220 assigns the VCs by programming a classifier lookup table. For one embodiment, the switch 220 performs the selection according to a load balancing procedure. During this procedure, the switch 220 measures the amount of data transmitted by each queuing buffer set in a predetermined period of time and assigns the VC to the buffer set having the higher measured amount of transmitted data—i.e., the lower amount of data stored. For example, the switch 220, in trying to maintain the total bandwidth of all VCs assigned to each buffer set approximately equal, assigns the VC request to the buffer set having the lower total bandwidth. For alternative embodiments, alternate procedures may be used in the process of selecting a queuing buffer set, such as a round robin procedure, where even and odd numbered VCs are sent to separate queuing buffer sets, or a random selection. As a result of the operation performed at step 440, the VC is assigned either to buffer set 320, at step 450, or in buffer set 330, at step 455.

Figure 4b is a flowchart representing a detailed process of storing cells in designated buffer sets. As shown in Figure 4b, the classifier 310 receives the cells at step 460. At step 465, the classifier 310 consults the programmed classifier lookup table to determine the buffer set where the cells are to be stored. Based on the lookup table, at step 470, the classifier 310 selects a buffer set for the cells associated with each particular VC, storing the cells either in buffer set 320, at step 475, or in buffer set 330, at step 476.

Figure 4c is a flowchart representing a detailed process of selecting a communication link. For illustrative purposes, the flowchart shows the

selection of a communication link in a system including two links. The flowchart, however, may be applicable to systems containing multiple communication links.

As shown in the flowchart of Figure 4c, at step 480, the cells stored in buffer sets 320, 330 are retrieved. A selection of the link available to carry the cells is performed at step 485. Assuming both links are up and available, at step 494, the cells associated with VCs assigned to buffer set 320 are transmitted through multiplexing device 340 along link 260. At the same time, if both links are available, the cells associated with VCs assigned to buffer set 330 are transmitted along link 280 at step 496. In case of a link failure, the cells stored within both buffers sets 320, 330 are transmitted to the available link at step 492.

Referring back to Figure 3, the multiplexing device 340 allocates the cells stored within the queuing buffer sets 320, 330 to either link 260 or link 280. For this embodiment, when both links are available, the cells stored in buffer set 320 are selectively sent along link 260, while the cells stored in buffer set 330 are selectively sent along link 280. Multiplexing device 340, however, has the capability to switch the flow of cells from one link to the other based on information received from processor 350. The processor 350 is connected to both links 260, 280 and detects any failure on those links. If a failure is detected, the processor 350 accesses the multiplexing device 340 and reprograms it to switch the cells from the failing link. For example, if a failure is detected on link 260, the multiplexing device 340 switches the cells transmitted along link 260, previously stored in buffer set 320, to link 280. Therefore, following the switch, link 280 will carry cells coming from both queuing buffer sets 320 and 330. If a failure is detected on link 280, cells transmitted along link 280 are switched in a similar fashion to link 260, to be transmitted together with the cells already carried by link 260.

The receiver line card 360 shown in Figure 3 receives the cells transmitted along both links 260 and 280, or along any one of the links in case of failure of the other. The receiver line card 360 further comprises a multiplexing

device 370, a classifier device 380 coupled to the multiplexer 370, a queuing engine 390 containing multiple queues, and a processor 395 connected to both links 260, 280.

The multiplexing device 370 receives and merges the transmitted cells and forwards the merged cells to the classifier 380. The classifier 380 classifies the cells based on the VPI and VCI of each cell and selects a LCN that is inserted into the cell header in a LCN field and is used for further storing of the cell. The classifier 380 then sends the classified cells to the queuing engine 390.

The cells are classified irrespective of the link that carried them. The cells having the same VPI, VCI, and LCN are stored in the same queue within the queuing engine 390. Additionally, each queue stores the cells in the order they are received, thereby maintaining the order of the initial VCs of the communication. Processor 395 is provided to monitor both links 260 and 280 and to detect any link failure that cannot be detected by processor 350 within the transmitter line card 300. When a failure is detected, processor 395 signals the failure to processor 350 and prompts it to reprogram the multiplexing device 340.

Figure 5 illustrates another embodiment of the present invention. System 500, residing in an ATM network (not shown), comprises a transmitter switch 520 coupled to a receiver switch 540 via communication links 560 and 580. Data is received on input links 510 and 511 and is forwarded to a switching fabric 610 having multicast capability through input line cards 530 and 535, and further to transmitter line cards 600 and 620 using the multicast capability. The transmitter line cards 600, 620 are described in detail below. The transmitter line card 600 is coupled to communication link 560, while the transmitter line card 620 is coupled to link 580, both line cards facilitating the transmission of data along links 560, 580 to the receiver switch 540. The receiver switch 540 includes receiver line cards 640, 660, which will also be described in detail below. The receiver line card 640 is coupled to link 560, while receiver line card 660 is coupled to link 580, both line cards receiving data

transmitted along those links. Finally, within the receiver switch 540, data is transferred from the receiver line cards 640, 660 to output line cards 550 and 570 and further to output links 590, 591. The exemplary environment illustrated in Figure 5 is not intended to limit the present invention. Alternative hardware environments may be used without departing from the scope of the present invention. The transmitter line cards 600, 620 and the receiver line cards 640, 660 will now be described in further detail.

As illustrated in Figure 6, the transmitter line cards 600, 620 are similar in architecture and functionality to the transmitter line card 300 shown in Figure 3 and described in detail above. Transmitter line cards 600, 620 include (1) classifier devices 601 and 621 respectively, for classifying the information cells; (2) queuing buffer sets 602, 603, and 622, 623 respectively, for storing the classified cells in multiple queues; and (3) multiplexing devices 604 and 624 respectively, for allocating the cells stored within the buffers to the respective links.

Referring also to Figure 6, the receiver line cards 640, 660 are similar in architecture and functionality to the receiver line card 360 shown in Figure 3 and described in detail above. Receiver line card 640 is coupled to link 560 and receives cells transmitted along link 560, while the receiver line card 660 is coupled to link 580 and receives cells transmitted along link 580. Similar to line card 360 of Figure 3, receiver line cards 640, 660 include (1) multiplexing devices 641 and 661 respectively, for receiving and merging the transmitted cells; (2) classifier devices 642 and 662 respectively, for classifying the cells sent by the multiplexing devices; (3) queuing engines 643 and 663 respectively, for storing in multiple queues the cells forwarded by the classifier devices; and (4) processors 645 and 665 respectively for monitoring the receiver line cards 640, 660 and their respective links.

The switching fabric 610 is provided to receive data from input links 510, 511, and to forward the information to both transmitter line cards 600, 620. For one embodiment, the multicast capability of the switching fabric 610 duplicates

data received from links 510, 511, forwarding the same information to each line card 600, 620. When both links are active, the multiplexing device 604 within line card 600 is coupled to link 560 and forwards the cells stored in buffer set 602 to link 560 for transmission. At the same time, the multiplexing device 604 routes the cells stored in buffer set 603 to a disposal bin (not shown). The multiplexing device 624 within line card 620 is coupled to link 580 and forwards the cells stored in buffer set 623 to link 580 for transmission, while sending the cells stored in buffer set 622 to the disposal bin. Because data is duplicated, the cells stored in buffer sets 602 and 622, as well as the cells stored in buffer sets 603 and 623, are identical. Therefore, assuming that both links 560 and 580 are functional, entire data received along links 510, 511 is carried and delivered to the receiver line cards 640 and 660.

The multiplexing devices 604, 624 can switch the flow of cells based on information received from processors 605 and 625 respectively. Processor 605 communicates with processor 625, both being designed to monitor their respective transmitter line cards and respective links. For example, if a failure is detected on link 560 or within the transmitter line card 600, processor 605 sends a signal to processor 625, which reprograms the multiplexing device 624 to switch the cells stored in buffer set 622 from the disposal bin to link 580. Similarly, if a failure is detected on link 580 or within the transmitter line card 620, processor 625 instructs processor 605 to switch the cells stored in buffer set 603 from the disposal bin to link 560.

On the receiving end, processor 645 within the receiver line card 640 communicates with processor 665 within line card 660 and with processor 605 within the transmitter line card 600, while processor 665 is connected to processor 625 within transmitter line card 620. If a failure is detected on link 560, processor 645 sends a signal to processor 665, instructing it to communicate with processor 625 to direct the multiplexing device 624 to switch the cells stored in buffer set 622 from the disposal bin to link 580. If a failure is detected within the receiver line card 640, in addition to the previous task, processor 645

also instructs processor 605 to reprogram the multiplexing device 604 to stop all cells from being transmitted along link 560, i.e. to switch cells stored in buffer set 602 from link 560 to the disposal bin. Conversely, failures on link 580 or within receiver line card 660 are treated in a similar fashion by the corresponding processors.

Once received and stored in the queues within the queuing engines 643 and 663, cells pertaining to the same VC are merged within the receiver switch 540 and forwarded to output line cards 550 and 570 shown in Figure 5. Each output line card 550, 570 includes a classifier device (not shown), which receives cells from both receiver line cards 640 and 660 and classifies them irrespective of their originating line card. Each output line card 550, 570 further includes a processor (not shown) connected to processors 645 and 665 within the receiver line cards 640, 660. Because the classifier devices classify the cells irrespective of their originating receiver line card, the functionality of the entire system 500 is not affected by the failure of any links or line cards. If, for example, a failure is detected within receiver line card 640, but cells are still transmitted along link 560, although processor 605 has been advised to stop traffic on link 560, processor 645 instructs the processors within the output line cards 550, 570 to reprogram the corresponding classifier devices to stop receiving cells from receiver line card 640 and to select line card 660 as supplier of cells. Any failure within any other line card can be similarly solved.

For another embodiment of the present invention, the switching fabric 610 does not replicate cells. The switching fabric 610 receives data from input links 510, 511, and forwards distinct data to each transmitter line card 600 and 620. The cells forwarded to line card 600 are classified and stored in buffer set 602 for transmission along link 560, while the cells forwarded to line card 620 are classified and stored in buffer set 623 for transmission along link 580. Buffer sets 603 and 622 remain empty and no cells are routed to the disposal bin. In the event of a link or card failure, input line cards 530 and 535 in Figure 5 can be instructed to reroute all traffic to the still functional card or link. For example,

if a failure is detected on link 560 or within the transmitter line card 600, the input line cards 530 and 535 request that the switching fabric 610 stop sending traffic to line card 600 and instead forward all data intended for line card 600 to line card 620. The switching fabric 610 may be provided with a lookup table to redirect the cells. Alternatively, the input line cards 530 and 535 may be made capable to change the destination card address of each cell intended for line card 600. As a result, data stored in both buffer sets 622 and 623 is transmitted along link 580. All other operations involving other components of the system 500 are similar to the ones discussed above.

In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.